DOCUMENT RESUME

ED 110 505 TM 004 778

AUTHOR Epstein, Marion G.
TITLE Classification Scheme for Items in CAAT.
INSTITUTION Educational Testing Service, Princeton, N.J.
PUB DATE [Oct 74]
NOTE 10p.; Paper presented at the Computer Assisted Test Construction Conference (San Diego, California, October 14-16, 1974); For a related document, see TM 004 776

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS *Classification; Codification; Computer Oriented Programs; *Computer Programs; *Item Banks; *Test Construction; *Test Selection

ABSTRACT

In planning the development of the system for computer assisted assembly of tests, it was agreed at the outset that one of the basic requirements for the successful initiation of any such system would be the development of a detailed item content classification system. The design of the system for classifying item content is a key element in systems development. Without an adequate classification system, it would be impossible to write test specifications and assemble tests that sample required content areas consistently and without undesirable redundancy. This system has a unique identification number for each item and its classification, the computerized file contains a complete history of the previous uses of the item, up to five sets of statistics including difficulty level and discrimination index, codes for the security level and present activity status of the item and up to 15 12-letter key words--which in the case of some verbal items constitute the complete item. Whether the whole items are stored in the computer or remain in a card file for manual retrieval, the selection process of surveying the pool of eligible items and selecting appropriate ones is the same. (Author/BJG)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

CATC Conference
San Diego, California
October 14-16, 1974

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED110505

CLASSIFICATION SCHEME FOR ITEMS IN CAAT

Marion G. Epstein
Director, Development and Analysis
CEEB Division
Educational Testing Service

In planning the development of the system for computer assisted assembly of tests, it was agreed at the outset that there were two basic requirements for the successful initiation of any such system - the development of detailed item content classification systems and the spelling out of the professional judgments made in building a test from a pool of items to fit detailed content, ability and statistical specifications in terms precise enough to be translated into computer programs. One guiding principle has been that the computer is to serve the professional staff, not supersede them. A second principle was that the format for item classification should be flexible enough to serve a variety of subject areas even though the methods of classification might be radically different.

Certainly one of the most important aspects of the development of a system for computer assisted test assembly is the design of the system for classifying item content. Without an adequate classification system, it would be impossible to write test specifications and assemble tests that sample required content areas consistently and without undesirable redundancy. However, for the selection of items from a large pool for

TM004778

secure national tests, such as the Scholastic Aptitude Test, much more information is needed. In addition to a unique identification number for each item and its classification, the computerized file contains a complete history of the previous uses of the item, up to five sets of statistics including difficulty level and discrimination index, codes for the security level and present activity status of the item (i.e., whether it is in an active or recently retired test or is "free use"), and up to fifteen twelve-letter key words - which, in the case of some verbal items constitute the complete item. Whether the whole items are stored in the computer or remain in a card file for manual retrieval, the selection process of surveying the pool of eligible items and selecting appropriate ones is the same. Incidentally, our current computerized pool contains more than 15,000 verbal aptitude items, about 12,000 mathematics items and several hundred American History items.

The format for content classification allows for any number of four-letter major categories, and within each of these categories, nine digits, each with ten levels - a staggering number of possible unique classifications. To date, we have used this basic format for very different approaches to classification of verbal items for the SAT, for the whole domain of mathematics content from elementary level through calculus and for the social sciences.

In developing subject-matter classifications, we have tried to make them detailed enough to be useful for the assembly of unit and

end-of-course tests as well as broad survey-type tests such as the College Board Achievement Examinations. In developing ability classifications in each subject area, we have tried to reflect the objectives of education in the field in terms of cognitive levels or thought processes. Although we have not yet tried it, we think we can use a combination of the content and ability dimensions to match individual school's behavioural objectives should we move to assembling tests on request for this purpose.

In the verbal area, the major categories are acronyms for item types such as ANAL for analogies, ANTM for antonyms or RC for reading comprehension. In mathematics, on the other hand, the major categories are broad content areas such as ALGB for algebra or ARIT for arithmetic or CALC for calculus. In the social sciences, we have used major divisions or disciplines such as GEOG for geography or AMHY for American History. The digits are used for finer distinctions of content or item type within each major category and for such information as lowest appropriate grade level, type of setting, and a taxonomy of ability or type of thinking required. It is, of course, essential that a group developing and using such a system agree on the meaning of each classification and use it consistently, since the computer couldn't care less what the letters and numbers stand for. And therein lies part of the problem. When you are testing skills rather than specific content, item types can be used and there is no overlap across major categories.

When, as in mathematics, you are dealing with specific content, many of the best questions cross major categories and could be classified in more than one way. There must then be clearly stated agreement on where to place certain problems. For example, we have agreed that problems involving geometric concepts which require algebra in their solution will be classified under geometry. Key words are used to control the number of questions in a test which call on the same techniques or concepts in their solution. For example, many questions may require the use of averages - although averages are explicitly included under arithmetic, geometry or algebra questions which use averages in their solution will have the word average listed under the key words and we can ask the computer to limit the number of such questions selected.

The verbal classification is simplest and uses a maximum of six digits. Here's an example (Example 1) This is an analogy item in which is science oriented, represents a concrete analogy (stem and choices represent real things) and is overlapping (some aspects of stem reappear in options. The "grade level" is appropriate for SAT (11th - 12th grade).

In mathematics, in each major category, the first digit, which we refer to as a major-minor category, is used for up to ten large subdivisions of the major category, such as algebraic operations or linear functions in algebra. (I have the complete classification with me if anyone is interested.) The second and third digits refer

back to the first digit, breaking it down further providing theoretically, if not practically, up to one hundred subdivisions of each major-minor category and one thousand unique classifications within each major area. The fourth and fifth digits are not currently used, but are available if finer classifications should be needed.

Digits 6-9 have been assigned the same meaning over all mathematics questions

Digit 6 controls for curriculum with four levels of traditional versus modern content and symbolism

Digit 7 indicates grade level from 1 for primary to 9 for advanced college and graduate level

Digit 8 describes the nature of the setting to insure variety in a test

Digit 9 we call the ability-process dimension - a hierarchial classification into six levels which is our pragmatic adaptation of the Bloom taxonomy

For example (Example 2) This classification describes an arithmetic question testing properties of integers; dealing with powers or roots of odd and even numbers. The content is modern; grade level junior high school; setting abstract with numbers only; ability level - comprehension of a concept

In social studies, the use and pairing of digits varies with the major category except that digits 7-9 are used across all areas for grade levels, type of stimulus, and ability tested. For example (Example 3)

This would be an item in Western European History. Chronologically, it is 16th century; geographically France; its subject matter is Governmental Institutions; its emphasis is Foreign Affairs, specifically in relation to Russia. The grade level is college; it is based on a quotation and the ability level is application.

This gives you, I hope, some flavor of our classification approaches. For specific courses and specific texts for classroom testing, a classification system could be much simpler and more limited, but some classification system is absolutely essential if tests are to be selected from a pool to serve their intended purpose well.

EXAMPLE 1

				1	2	3	4	5	6	7	8	9
A	N	A	L	3	1	2	Ø	6	Ø	Ø	Ø	Ø

SNAKE:PYTHON:

(A) BIRD:STARLING (B) FLOWER:BLOSSOM (C) MAMMAL:REPTILE

(D) LION:TIGER (E) RAT:MOUSE

THIS IS AN ANALOGY ITEM WHICH IS:

SCIENCE ORIENTED (1st digit 3)

CONCRETE (stem and choices all represent real things) (2nd digit 1)

OVERLAPPING (some aspects of stem appears in options) (3rd digit 2)

GRADE LEVEL - appropriate for SAT or above (5th digit 6)

4th digit is always zero for analogies

EXAMPLE 2

				1	2	3	4	5	6	7	8	9
A	R	I	T	1	2	5	Ø	Ø	2	4	5	3

AN ARITHMETIC QUESTION WHICH DEALS WITH

PROPERTIES OF INTEGERS (1st digit 1)

ODD AND EVEN NUMBERS (2nd digit 2)

POWERS OR ROOTS (3rd digit 3)

MODERN MATH, NEUTRAL SYMBOLISM OR LANGUAGE (6th digit 2)

GRADE LEVEL - 7-9 (7th digit 4)

ABSTRACT SETTING - NUMBERS ONLY (8th digit 5)

ABILITY LEVEL - COMPREHENSION (9th digit 3)

EXAMPLE 3

				1	2	3	4	5	6	7	8	9
W	E	H	Y	4	2	3	3	2	4	4	4	4

THIS CLASSIFICATION DESCRIBES AN ITEM IN THE HISTORY OF WESTERN EUROPE WHICH DEALS WITH

THE PERIOD 1600 - 1715 (1st two digits 42)

FRANCE (3rd digit 3)

GOVERNMENTAL INSTITUTIONS (4th digit 4)

FOREIGN AFFAIRS, SPECIFICALLY (5th digit 2)

WITH RUSSIA (6th digit 4)

COLLEGE-LEVEL MATERIAL (7th digit 4)

BASED ON A QUOTATION (8th digit 4)

ABILITY TESTED IS APPLICATION (9th digit 4)